

jc997 U.S. PTO
09/960731
09/24/01

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Patent Application of:

Achim Mueller

Application No.: Group Art Unit: (unassigned)

Filed: (concurrently) Examiner: (unassigned)

For: METHODS FOR SPEECH PROCESSING

**SUBMISSION OF CERTIFIED COPY OF PRIOR FOREIGN
APPLICATION IN ACCORDANCE
WITH THE REQUIREMENTS OF 37 C.F.R. §1.55**

Assistant Commissioner for Patents
Washington, D.C. 20231

Sir:

In accordance with the provisions of 37 C.F.R. §1.55, the applicant(s) submit(s) herewith a certified copy of the following foreign application:

German Patent Application No. 100 47 172.2

Filed: 22 September 2000

It is respectfully requested that the applicant(s) be given the benefit of the foreign filing date(s) as evidenced by the certified papers attached hereto, in accordance with the requirements of 35 U.S.C. §119.

Respectfully submitted,

STAAS & HALSEY LLP

Date: 9/24/01

By: Richard A. Gollhofer
Richard A. Gollhofer
Registration No. 31,106

700 11th Street, N.W., Ste. 500
Washington, D.C. 20001
(202) 434-1500

This Page Blank (uspto)

ATTACHMENT 1(e)

EXPLANATIONS OF RELEVANCY OF REFERENCES	ATTORNEY DOCKET NO.	APPLICATION NO.
	1454.1096	
	FIRST NAMED INVENTOR	
	Achim Mueller	
FILING DATE (Concurrently)	GROUP ART UNIT	

The relevancy of German Patents 19636739 C1 and 19719n381 C1 is discussed in the application.

jc997 U.S. PTO
09/960731

09/24/01

This Page Blank (uspto)

BUNDESREPUBLIK DEUTSCHLAND



JC997 U.S. PTO
09/960731
09/24/01

Prioritätsbescheinigung über die Einreichung einer Patentanmeldung

Aktenzeichen: 100 47 172.2
Anmeldetag: 22. September 2000
Anmelder/Inhaber: Siemens Aktiengesellschaft,
München/DE
Bezeichnung: Verfahren zur Sprachverarbeitung
IPC: G 10 L 15/00

CERTIFIED COPY OF
PRIORITY DOCUMENT
CERTIFIED COPY OF
PRIORITY DOCUMENT

Die angehefteten Stücke sind eine richtige und genaue Wiedergabe der ursprünglichen Unterlagen dieser Patentanmeldung.

München, den 29. März 2001
Deutsches Patent- und Markenamt
Der Präsident
Im Auftrag

This Page Blank (us)

Beschreibung

Verfahren zur Sprachverarbeitung

5 Die Erfindung betrifft Verfahren, eine Anordnung und ein Computerprogrammprodukt zur Sprachverarbeitung.

Verfahren und Anordnungen zur Sprachverarbeitung sind etwa aus US 6 029 135, US 5 732 388, DE 19636739 C1 und DE 10 19719381 C1 bekannt. Insbesondere die Realisierung multilingualer und sprachraumunabhängiger Sprachsynthesesysteme basiert zu einem großen Teil auf datengetriebenen Modulen. Diese Module, zum Beispiel Prosodiegenerierungsmodule, verwenden in der Regel Lernverfahren. Die Lernverfahren können im Allgemeinen gut für mehrere Sprachen und Anwendungen eingesetzt werden. Jedoch müssen oft die Eingangsgrößen mühsam per Hand optimiert werden.

Für den Fall der symbolischen Prosodie, also insbesondere der Phrasengrenzenvorhersage und der Vorhersage akzentuierter Wörter z.B. durch entsprechende Grundfrequenzerzeugung, sind die folgenden Lerntechniken angewandt worden: Für die Phrasengrenzvorhersage Ansätze, die auf Klassifikations- und Regressionsbäumen (CARTs) basieren von Julia Hirschberg und Pillar Prieto: „Training Intonational Phrasing Rules Automatically for English and Spanish Text-to-speech“, Speech Communication, 18, S. 281-290, 1996, und Michell Q. Wang und Julia Hirschberg: „Automatic Classification of Intonational Phrasing Boundaries“ Computer Speech and Language, 6, S. 175-196, 1992, Ansätze, die auf Hidden-Markov-Modellen (HMM) basieren von Alan W. Black und Paul Taylor: „Assigning Phrase Breaks from Part-of-Speech Sequences“, Eurospeech, 1997, und Ansätze, die auf Neuronalen Netzen basieren von Achim F. Müller, Hans Georg Zimmermann und Ralf Neuneier: „Robust Generation of Symbolic Prosody by a Neural Classifier Based on autoassociators“, ICASSP, 2000. Für die Vorhersage von Akzenten oder akzentuierten Wörtern wurden CARTs verwendet von Julia

Hirschberg: „Pitch Accent in Context: Predicting Prominence from Text“, Artificial Intelligence, 63, S. 305-340, 1993, wohingegen Neuronale Netze eingesetzt wurden von Christina Widera, Thomas Portele und Maria Wolters: „Prediction of Word

5 Prominence“, Eurospeech 1997. Eine Interpretation des Einflusses der verwendeten Eingangsgrößen ist dabei in der Regel nicht möglich. Dies gilt insbesondere für neuronale Netze.

Für den Fall der Grundfrequenzerzeugung (f0-Generierung) ist dieses Problem auch bekannt. So werden zum Beispiel in Gerit 10 P. Sonntag, Thomas Portele und Barbara Heruft: „Prosody Generation with a Neural Network: Weighing the Importance of Input Parameters“, ICASSP, 1997, die Eingangsgrößen heuristisch optimiert.

15 Davon ausgehend liegt der Erfindung die Aufgabe zugrunde, Sprachverarbeitungsverfahren dadurch zu verbessern, dass bei der Abbildung von Sprachmerkmale enthaltenden Eingangsgrößen auf Ausgangsgrößen eine höhere Berücksichtigung der wichtigen Eingangsgrößen erfolgt. Weiterhin sollen ein Verfahren, eine 20 Anordnung und ein Computerprogrammprodukt angegeben werden, bei denen sich die Abbildung der Eingangsgrößen auf die Ausgangsgrößen genauer und schneller ermitteln lässt.

25 Diese Aufgabe wird durch Verfahren, eine Anordnung und ein Computerprogrammprodukt mit den Merkmalen der unabhängigen Ansprüche gelöst.

30 Dabei werden die Eingangsgrößen mit unterschiedlichen Gewichten auf die erzeugten Ausgangsgrößen abgebildet. Durch die Gewichte kann die Wichtigkeit einzelner Eingangsgrößen berücksichtigt werden. Die Gewichte lassen sich dabei in beliebiger Form ausbilden, so zum Beispiel durch Multiplikation mit einem Faktor, durch Addition eines Summanden oder durch 35 eine beliebige Funktion, die bei Anwendung auf die Eingangsgröße diese entsprechend abändert.

Gemäß der Erfindung werden die Gewichte nicht mehr heuristisch gefunden, sondern die Abbildung der erzeugten Ausgangsgröße (Ist-Zustand) wird mit der Abbildung der zu erzeugenden Ausgangsgröße (Soll-Zustand) verglichen. Daraus wird eine Änderungsvorschrift für die Abbildung berechnet, wobei diese Änderungsvorschrift mit der Maßgabe, das heißt ganz gezielt, so berechnet wird, dass die Gewichte von Eingangsgrößen verringert werden, die einen geringen Einfluss auf die Ausgangsgröße haben. Ein geringer Einfluss heißt, dass die Eingangsgröße wenig relevante Informationen trägt. Solch ein geringer Einfluss stellt sich zum Beispiel dadurch dar, dass sich bei einer großen Änderung der Eingangsgröße die Ausgangsgröße nur gering ändert, oder dass sich die Ausgangsgröße stark ändert, obwohl die Eingangsgröße konstant bleibt.

15

Insbesondere ist es vorteilhaft, die Verringerung der Gewichte in Abhängigkeit vom Wert anderer Gewichte vorzunehmen. Dies ergibt sich daraus, dass nicht die absolute Größe der Gewichte maßgebend ist, sondern die Gewichtung der Eingangsgrößen in Relation zueinander. Die anderen Gewichte können dabei selektiv oder vollständig berücksichtigt werden. Ein Aufsummieren über die anderen Gewichten ist dabei ebenso denkbar wie eine Mittelwertbildung. Je nach Ausprägung der Gewichte können auch nur deren Beträge berücksichtigt werden, zum Beispiel indem diese quadriert werden.

25

Die Verringerung der Gewichte lässt sich in einfacher Weise mit einer vorgebbaren Verringerungsrate steuern. Es ist deshalb vorteilhaft, eine solche in das Verfahren aufzunehmen.

30

Enthält die Abbildung mehrere Abbildungsschichten, zum Beispiel in Form von mehreren miteinander verknüpften Funktionen oder auch in Form von Schichten eines Neuronalen Netzes, so wird die Gewichtung vorzugsweise bei der ersten Abbildungsschicht vorgenommen. Dadurch erhält man die am besten interpretierbaren Ergebnisse.

Insbesondere wird die Gewichtung einer Eingangsgröße vorgenommen, bevor sie mit einer anderen Eingangsgröße verrechnet wird. Um den Rechenbedarf zu verringern, können Gewichte, die unterhalb eines bestimmten Schwellwertes liegen, 5 auf Null gesetzt werden. Das heißt, dass die zugehörige Eingangsgröße in der Abbildung nicht mehr berücksichtigt wird.

Sogenannte Ausreißer bei den Eingangsgrößen, die besonderes hohe Werte aufweisen, lassen sich dadurch unterdrücken, dass 10 die Abbildung für die entsprechende Eingangsgröße eine Übertragungsfunktion aufweist, die für betragsmäßig große Werte der Eingangsgröße eine geringe Steigung aufweist.

Damit die Ausreißer nur einen geringen Einfluss auf die restliche Abbildung haben, wird diese Übertragungsfunktion bevorzugt als erstes auf die Eingangsgrößen angewandt. Das heißt, noch bevor die restlichen Teile der Abbildung angewandt werden. Als Übertragungsfunktion eignet sich insbesondere eine sigmoide Übertragungsfunktion, wie zum Beispiel der Tangens-20 hyperbolicus oder die logistische Funktion.

Die Abbildung lässt sich ganz oder teilweise als Neuronales Netz realisieren. Dabei werden die Eingangsgrößen über künstliche Neuronen mit der mindestens einen durch die Abbildung 25 erzeugten Ausgangsgröße verknüpft. In dieser Ausgestaltung können die Gewichte mit den Gewichten des neuronalen Netzes identifiziert werden.

Durch die Verringerung von Gewichten von Eingangsgrößen, die 30 einen geringen Einfluss auf die Ausgangsgrößen haben, wird insbesondere die Generalisierungsfähigkeit eines Neuronalen Netzes verbessert.

Wird das neuronale Netz in einem Lernverfahren trainiert, so 35 werden vorzugsweise der Vergleich der erzeugten Ausgangsgröße mit der zu erzeugenden Ausgangsgröße und die Berechnung der Änderungsvorschrift aus dem Vergleich mit der Maßgabe, dass

die Gewichte von Eingangsgrößen verringert werden, die einen geringen Einfluss auf die Ausgangsgröße haben, wiederholt ausgeführt, so dass man iterativ zu einer Abbildung gelangt, bei der die erzeugten Ausgangsgrößen den zu erzeugenden Aus-
5 gangsgrößen immer mehr entsprechen und die Gewichte derjenigen Eingangsgrößen weiter verringert werden, die einen gerin-
gen Einfluss auf die Ausgangsgröße haben.

10 Besonders bevorzugt wird dieses Verfahren mit mehreren Sätzen von Sprachmerkmale enthaltenden Eingangsgrößen und diesen Sätzen jeweils zugeordneten Sätzen mindestens einer zu erzeugenden Ausgangsgröße durchgeführt. Hierdurch kann die Abbil-
dung weiter verfeinert bzw. das Neuronale Netz besser trai-
niert werden.

15

15 Liegen mehrere Sätze von Eingangsgrößen mit jeweils gleichartigen Eingangsgrößen, mindestens eine zu erzeugenden Aus-
gangsgröße und entsprechend mehrere Abbildungen vor, so lässt sich das Verfahren in erfindungsgemäßer Fortbildung dadurch
20 verbessern, dass die Änderungsvorschriften für die einzelnen Abbildungen so berechnet werden, dass sich dieselben Abbil-
dungen für unterschiedliche Sätze von Eingangsgrößen ergeben. Dies röhrt daher, dass bei Sätzen gleichartiger Eingangsgrö-
ßen die Abbildungen auf die mindestens eine durch die oder
25 eine der Abbildungen zu erzeugende Ausgangsgröße identisch sein müssen.

25 Dies kann zum Beispiel dadurch erreicht werden, dass alle Ab-
bildung identisch oder nahezu identisch initialisiert wer-
den und danach auf alle Abbildungen nur noch identische Ände-
rungsvorschriften angewandt werden.

30 Ausgehend von identisch initialisierten Änderungsvorschriften ergeben sich identische Änderungsvorschriften für alle Abbil-
dungen zum Beispiel dadurch, dass zunächst vorläufige Ände-
rungsvorschriften berechnet werden, bei deren Berechnung nur
35 ein einzelner Satz von Eingangsgrößen, die zugeordnete Abbil-

dung, die erzeugte Ausgangsgröße und die zu erzeugende Aus-
gangsgröße berücksichtigt wird. Diese vorläufigen Änderungs-
vorschriften berechnet man für alle vorgegebenen Sätze. Da-
nach werden die Mittelwerte der vorläufigen Änderungsvor-
5 schriften ermittelt. Diese ergeben die Änderungsvorschriften,
mit denen dann die Abbildungen tatsächlich geändert werden.

Auch bei diesem Verfahren lassen sich vorteilhaft Übertra-
gungsfunktionen zum Dämpfen von Ausreißern und Neuronale Net-
10 ze mit künstlichen Neuronen und Gewichten einsetzen. Diese
Ausgestaltungen ergeben sich aus den Unteransprüchen.

Um Speicherplatz zu sparen, werden die identischen Abbildun-
gen vorzugsweise am gleichen Speicherort gespeichert. Darüber
15 hinaus können die Abbildungen einen oder mehrere gemeinsame
Abbildungsteile aufweisen.

Weiterhin können die Schritte des Vergleichens der von den
Abbildungen erzeugten Ausgangsgrößen mit den von den Abbil-
20 dungen zu erzeugenden Ausgangsgrößen und die Schritte des Be-
rechnens der Änderungsvorschriften für die Abbildungen aus
dem Vergleich, so dass sich dieselben Abbildungen für unter-
schiedliche Sätze von Eingangsgrößen ergeben, wiederholt wer-
den. So lässt sich iterativ die Abbildung mit dem kleinsten
25 Fehler bestimmen.

Ziel der Erfindung ist es nicht nur, eine oder mehrere Abbil-
dungen in der beschriebenen Weise zu berechnen, sondern die
Abbildungen darüber hinaus auch auf Eingangsgrößen anzuwen-
30 den, von denen die durch die Abbildung zu erzeugenden Aus-
gangsgrößen nicht bekannt sind. Dazu wird ein Verfahren ange-
wandt, bei dem eine in der zuvor geschilderten Weise erzeugte
Abbildung verwendet wird.

35 Eine Anordnung, die eingerichtet ist, eines der geschilderten
Verfahren auszuführen, lässt sich zum Beispiel durch entspre-

chendes Programmieren eines Computers oder einer Rechenanlage realisieren.

Ein Computerprogrammprodukt, das Softwarecodeabschnitte enthält, mit denen eines der geschilderten Verfahren auf einem Computer ausgeführt werden kann, lässt sich durch geeignete Implementierung des Verfahrens in einer Programmiersprache ausführen. Die Softwarecodeabschnitte werden dazu gespeichert. Dabei wird unter einem Computerprogrammprodukt das Computerprogramm als handelbares Produkt verstanden. Es kann in beliebiger Form vorliegen, so zum Beispiel auf Papier, einem Computer lesbaren Datenträgen oder über ein Netz verteilt.

Weitere wesentliche Merkmale und Vorteile der Erfindung ergeben sich aus der Beschreibung eines Ausführungsbeispiels anhand der Zeichnung. Dabei zeigt

Figur 1 den schematischen Aufbau eines Verfahrens zur Sprachverarbeitung;

Figur 2 den schematischen Aufbau eines Verfahrens zur Sprachverarbeitung mit mehreren Sätzen von Eingangsgrößen und

Figur 3 die gemäß dem Verfahren zur Sprachverarbeitung mit mehreren Sätzen von Eingangsgrößen berechneten Gewichte.

In Figur 1 erkennt man die Architektur des Verfahrens zur Sprachverarbeitung. Dabei werden 1 Sprachmerkmale enthaltenden Eingangsgrößen x_i in Form eines Eingangsvektors x über einen Eingang 1 dem Verfahren zugeführt. Über eine Vorverarbeitungsschicht 2 werden die Eingangsgrößen gewichtet und auf transformierte Eingangsgrößen x_i' in Form eines Vektors abgebildet. Die transformierten Eingangsgrößen x_i' werden am Ausgang 3 der Vorverarbeitungsschicht 2 ausgegeben und einem Au-

toassziator-Klassifikator-Netzwerk 4 zugeführt. Der Aufbau eines solchen Autoassziator-Klassifikator-Netzwerks ist zum Beispiel in Achim F. Müller, Hans Georg Zimmermann und Ralf Neuneier: „Robust Generation of Symbolic Prosody by a Neural 5 Classifier Based on autoassociators“, ICASSP, 2000, beschrieben. Durch das Autoassziator-Klassifikator-Netzwerk 4 werden die transformierten Eingangsgrößen x'_i und damit auch die Eingangsgrößen x_i auf die erzeugten Ausgangsgrößen y_i abgebildet. Die Ausgangsgrößen y_i werden schließlich am Ausgang 10 ausgegeben.

Im in Figur 1 dargestellten Ausführungsbeispiel enthält die Abbildung 2, 4 somit eine Vorverarbeitungsstufe 2 und ein Autoassziator-Klassifikator-Netzwerk 4. Alternativ können die 15 der Vorverarbeitungsstufe 2 zugewiesenen und im Folgenden beschriebenen Aufgaben aber auch dem Autoassziator-Klassifikator-Netzwerk 4 zugewiesen werden. Dieses muss auch nicht als Autoassziator-Klassifikator-Netzwerk 4 ausgebildet sein, sondern es sind je nach dem durch die Sprachverarbei- 20 tung zu lösenden Problem auch andere Neuronale Netze mit abweichenden Netzwerkarchitekturen verwendbar. Neben Neuronalen Netzen können darüber hinaus auch weitere geeignete Abbildungen eingesetzt werden, insbesondere wenn diese durch Lernverfahren trainierbar sind. Um herauszufinden, welche der 1 Eingangsgrößen x_i für die spezifische Aufgabenstellung in der 25 Sprachverarbeitung wichtig ist, ist die Vorverarbeitungsschicht 2 zwischen dem Eingang 1 und dem Autoassziator-Klassifikator-Netzwerk 4 angeordnet. Die 1 Eingangsgrößen x_i werden in dieser Vorverarbeitungsschicht 2 über eine Diago- 30 nalmatrix $w_{diag} = diag(w_1 \dots w_1)$ übertragen, so dass man am Ausgang 3 der Vorverarbeitungsschicht 2 die transformierten Eingangsgrößen x'_i erhält.

Die Verringerung der Gewichte wird im dargestellten Ausführungsbeispiel nur auf die Gewichte der Diagonalmatrix w_{diag} 35 angewandt. Dazu wird für die Neuronen der Vorverarbeitungs-

schicht 2 die Identitätsfunktion oder der Tangenshyperbolicus als Aktivierungsfunktion gewählt.

Zu Beginn der Trainingsphase werden alle Elemente der diagonalen Matrix w_{diag} , das heißt alle Gewichte w_i , mit 1 initialisiert. Somit werden die Eingangsgrößen x_i ohne Modifikation zum Autoassoziator-Klassifikator-Netzwerk 4 übertragen.

Zur Verringerung der Gewichte w_i der Eingangsgrößen x_i , die einen geringen Einfluss auf die Ausgangsgrößen y_i haben, wird nunmehr das im Folgenden beschriebene Verfahren angewandt. Zu einer Fehlerfunktion $F(w)$ für die Vorverarbeitungsschicht 2 wird ein Strafterm $P(w)$ hinzu addiert:

$$15 \quad \tilde{F}(w) = F(w) + \lambda \cdot P(w)$$

Dabei lässt sich über die vorgebbare Verringerungsrate λ der Einfluss des Strafterms $P(w)$ einstellen. Eine Möglichkeit der Wahl des Strafterms $P(w)$ ist $P(w) = \sum_k w_k^2$, mit $k = 1 \dots l$.
20 Damit wird die um den Strafterm $P(w)$ erweiterte Fehlerfunktion $\tilde{F}(w)$ zu

$$\tilde{F}(w) = F(w) + \lambda \sum_k w_k^2$$

25 Während der Lernphase werden die Gewichte zu jedem Iterationsschritt j unter Verwendung des Gradientenabstiegsverfahrens auf der Basis dieser erweiterten Fehlerfunktion trainiert:

$$30 \quad \begin{aligned} \tilde{w}^{j+1} &= \tilde{w}^j - \eta \nabla \tilde{F}(w) \\ &= \tilde{w}^j - \nabla \left[\eta F(w) + \eta \lambda \sum_k w_k^2 \right]. \end{aligned}$$

Dies ist die Änderungsvorschrift für die Gewichte in der Vorverarbeitungsstufe. Der Parameter η wird üblicherweise als

Lernrate bezeichnet und steuert die Schrittweite, die bei der Anpassung der Gewichte verwendet wird. Vorzugsweise wird die Lernrate η und die Verringerungsrate λ in allen Schritten konstant gehalten.

5

Es hat sich als wichtig herausgestellt, die Verringerungsrate λ sorgfältig zu wählen. Die Verringerungsrate λ sollte üblicherweise so klein wie möglich gewählt werden. Dadurch ist der Einfluss der Lernrate η in der Änderungsvorschrift, die 10 auf die Gewichte in der Vorverarbeitungsstufe 2 angewandt wird, größer als der Einfluss der Verringerungsrate λ . So können nichtlineare Beziehungen erfasst werden, die in den Daten verborgen sind. Auf der anderen Seite sollte die Verringerungsrate λ groß genug sein, so dass sie die Gewichte w_i 15 in der Diagonalmatrix w_{diag} der Vorverarbeitungsstufe 2 beeinflusst.

Nach mehreren Trainingsepochen und Anwendungen der Änderungsvorschrift auf die Gewichte w_i kann man das folgende Verhalten beobachten: Für einige Gewichte ist der Einfluss der Lernrate η größer als der Einfluss der Verringerungsrate λ . Für andere Gewichte ist jedoch der Einfluss der Verringerungsrate größer als der Einfluss der Lernrate η . Durch die richtige Wahl des Verhältnisses der Verringerungsrate λ zur 25 Verringerungsrate η können einige Gewichte auf oder nahezu auf Null verringert werden, während andere Gewichte eine nicht zu vernachlässigende Größe behalten. Die Gewichte nahe Null oder unter einem bestimmten Schwellwert werden als weniger wichtig für den Trainingserfolg des Autoassoziator-Klassifikator-Netzwerks 4 angesehen. Alle Gewichte des Autoassoziator-Klassifikator-Netzwerks werden ohne einen Strafterm $P(w)$ zur gleichen Zeit wie die Gewichte in der Vorverarbeitungsstufe 2 trainiert.

35 Das Konzept des Hinzufügens einer Vorverarbeitungsstufe 2, die den Eingang 1 über den Ausgang 3 der Vorverarbeitungsstufe 2 mit dem neuronalen Autoassoziator-Klassifikator-Netzwerk

4 verbindet, wird durch das geschilderte Verfahren auf die Analyse von in Wortkategoriefolgen (Part-of-Speech-Sequences) angeordneten Wortkategorien (Parts-of-Speech) als Eingangsgrößen x_i angewandt. Dadurch lässt sich der Einfluss bestimmter Wortkategorien auf die Phrasengrenzenvorhersage und/oder die Grundfrequenzerzeugung und insbesondere die notwendige Größe des Kontextfensters berechnen. Das Kontextfenster bestimmt, wie viele Sätze von Eingangsgrößen x_i in Form von Wortkategorien für die symbolische Prosodie berücksichtigt werden müssen. Ein Satz von Eingangsgrößen wird in diesem Fall durch alle Eingangsgrößen x_i gebildet, die zum gleichen Zeitpunkt t vorliegen. Eine Folge von Sätzen von Eingangsgrößen bildet damit eine Zeitreihe für diese Wortkategorien.

15 Figur 2 zeigt eine Architektur für das Verfahren zur Sprachverarbeitung, die vorteilhaft eingesetzt werden kann, wenn mehrere Sätze von Eingangsgrößen vorliegen, die jeweils gleichartige Eingangsgrößen enthalten. Dies ist zum Beispiel bei den gerade geschilderten Zeitreihen der Fall. Dabei wird 20 der Eingang 100-110 in Form von Eingangsclustern ausgebildet, so dass jedes Cluster für einen Satz von Eingangsgrößen zuständig ist. Im dargestellten Beispiel wird ein Satz von Eingangsgrößen durch die jeweils zu einem Zeitpunkt gehörenden Eingangsgrößen im Kontextfenster gebildet. Im Ausführungsbeispiel wurden für das Kontextfenster fünf Zeitschritte vor und 25 fünf Zeitschritte hinter dem aktuellen Zeitpunkt berücksichtigt. Jeder Satz von Eingangsgrößen enthält 35 Wortkategorien als Eingangsgrößen. Diese 35 Wortkategorien sind von Satz zu Satz gleichartig. Die Eingangsgrößen werden clusterweise über entsprechende Vorverarbeitungsschichten 200-210, die in ihrem Aufbau jeweils der Vorverarbeitungsschicht aus dem vorhergehenden Ausführungsbeispiel entsprechen, auf die Ausgänge 300-30 310 der Vorverarbeitungsschichten 200-210 übertragen. Von hier gelangen sie in ein gemeinsames Autoassoziator-Klassifikator-Netzwerk 400, aus dem sie beim Ausgang 500 ausgegeben werden. Das Autoassoziator-Klassifikator-Netzwerk 400

ist ein gemeinsamer und identischer Abbildungsteil aller Abbildungen.

Wesentlich ist nun, dass die in den Vorverarbeitungsschichten 5 200 bis 210 verwendeten Diagonalmatrizen von Gewichten nicht unabhängig voneinander berechnet werden. Vielmehr werden die Änderungsvorschriften für diese Matrizen so berechnet, dass sich dieselben Abbildungen für unterschiedliche, insbesondere alle, Sätze von Eingangsgrößen ergeben. Dazu werden zunächst 10 vorläufige Änderungsvorschriften für die in den Vorverarbeitungsschichten 200-210 angewandten Gewichte berechnet, indem die Änderungsvorschriften für jeden Satz von Eingangsgrößen und den jeweiligen in einer der Vorverarbeitungsschichten 15 200-210 angeordneten Teil der jeweils zugeordneten Abbildung unabhängig voneinander berechnet werden. Danach wird der Mittelwert der vorläufigen Änderungsvorschriften gebildet, der eine gemeinsame Änderungsvorschrift ergibt, die dann auf die Gewichte in allen Vorverarbeitungsschichten 200-210 angewandt wird. Sind die einander entsprechenden Gewichte in den 20 Vorverarbeitungsschichten 200-210 gleich initialisiert worden, so ergibt sich aus dieser Vorgehensweise, dass die Änderungsvorschriften für diese Gewichte so berechnet werden, dass sich dieselben Abbildungen für die unterschiedlichen Sätze von Eingangsgrößen ergeben.

25 Die Ergebnisse aus Experimenten zur Bestimmung der Größe des Kontextfensters für die Phrasengrenzenvorhersage sind in Figur 3 dargestellt. Die Figur 3 zeigt die Werte der Gewichte w_i in der diagonalen Matrix w_{diag} . Dabei werden pro Zeitschritt jeweils 35 Wortkategorien berücksichtigt. Die Anzahl der Zeitschritte beträgt gemäß den obigen Ausführungen 11. Der Mittelwert der Gewichte pro Zeitschritt ist jeweils durch einen horizontalen Balken dargestellt. Der Balken für das Zentrum des Phrasengrenzenkontextfensters ist fett dargestellt. Wie man erkennen kann, ist die Position rechts neben dem Zentrum des Phrasengrenzenkontextfensters, also die zeitlich direkt nach der Phrasengrenze liegende Position, dieje-

nige mit dem größeren Mittelwert. Das bedeutet, dass sie am wichtigsten für die Phrasengrenzenvorhersage ist. Weiterhin kann man sehen, dass die Positionen, die mehr als zwei Zeitschritte vom Zentrum entfernt sind, geringe Mittelwerte aufweisen und damit für die Phrasengrenzenvorhersage nicht relevant sind.

Dadurch dass die Gewichte von Eingangsgrößen, die einen geringen Einfluss auf die Ausgangsgröße haben, verringert werden, steigt insbesondere die Generalisierungsfähigkeit der verwendeten Abbildung und der häufig bei Neuronalen Netzen auftretende Effekt des reinen Auswendiglernens bleibt aus.

Die Anwendung einer in dieser Weise erstellten Abbildung auf vorherzusagende Phrasengrenzen liefert damit deutlich bessere Ergebnisse als Abbildungen nach dem Stand der Technik.

Das Verfahren nach Figur 1 kann auch zur Analyse des Einflusses der Eingangsgrößen bei der Grundfrequenzerzeugung eingesetzt werden. Für diese Einsatzzweck wird das Autoassoziator-Klassifikator-Netzwerk 4 in Figur 1 durch ein Neuronales Netz mit Standardarchitektur für die Grundfrequenzerzeugung der zu untersuchenden Sprache ersetzt. Ein solches Netz wird beispielsweise in Tao Jianhua, Cai Lianhong, Martin Holzapfel und Herbert Tropf: „A Neural Network based Prosodic Model of Mandarin TTS System“, ICSLPS, 2000, beschrieben.

Das Verfahren zur Analyse des Einflusses der Eingangsgrößen bei der Grundfrequenzerzeugung ist analog zum Verfahren für die Analyse der notwendigen Größe des Phrasengrenzenkontextfensters für die Anwendung in der symbolischen Prosodie. Die Eingangsgrößen repräsentieren allerdings in diesem Fall phonetische und linguistische Informationen. Diese sind im Gegensatz zu den Wortkategorien teilweise nicht symbolisch, sondern liegen in kontinuierlicher Form vor. Hierbei kann es vorkommen, dass einzelne Eingangsgrößen mit Ausreißern behaftet sind, das heißt, dass sie betragsmäßig sehr große Werte einnehmen können, die den Lernalgorithmus eines zu trainie-

renden Neuronalen Netzes stören. Um dies zu verhindern, wird vorzugsweise eine Übertragungsfunktion der Abbildung vorgeschoaltet, die solche Ausreißer dämpft. Dies lässt sich zum Beispiel dadurch realisieren, dass die Aktivierungsfunktion 5 in der Vorverarbeitungsschicht 2 als sigmoide Funktion gewählt wird.

Allen Ausführungsbeispielen liegt der Gedanke zugrunde, eine verbesserte Analyse von Sprache zu erzielen, indem unwichtige 10 Sprachmerkmale automatisch herausgefunden und in ihrem Einfluss auf die Vorhersage zurückgedrängt werden oder indem in Zeitreihen vorhandene Informationen gemeinsam ausgewertet werden. Dabei liegt es auch im Bereich der Erfindung, die dadurch gewonnenen Erkenntnisse bei der Sprachverarbeitung an- 15 zuwenden, indem die durch das Verfahren bzw. durch eine Anordnung, die eingerichtet ist, ein entsprechendes Verfahren auszuführen, gewonnene Abbildung bei der Sprachsynthese und/oder Spracherkennung verwendet wird.

Patentansprüche

1. Verfahren zur Sprachverarbeitung mit einer Abbildung (2, 4, 200-210, 400) von Sprachmerkmale enthaltenden Eingangsgrößen (x_i) auf mindestens eine durch die Abbildung (2, 4, 200-210, 400) erzeugte Ausgangsgröße (y_i) und mit mindestens einer durch die Abbildung (2, 4, 200-210, 400) zu erzeugenden Ausgangsgröße, bei dem

- a) die Eingangsgrößen (x_i) mit unterschiedlichen Gewichten (w_i) auf die durch die Abbildung (2, 4, 200-210, 400) erzeugte Ausgangsgröße (y_i) abgebildet werden,
- b) die von der Abbildung (2, 4, 200-210, 400) erzeugte Ausgangsgröße (y_i) mit der von der Abbildung (2, 4, 200-210, 400) zu erzeugenden Ausgangsgröße verglichen wird,
- c) aus dem Vergleich eine Änderungsvorschrift für die Abbildung (2, 4, 200-210, 400) berechnet wird und
- d) die Änderungsvorschrift zumindest teilweise mit der Maßgabe berechnet wird, dass die Gewichte (w_i) von Eingangsgrößen (x_i) verringert werden, die einen geringen Einfluss auf die Ausgangsgröße (y_i) haben.

2. Verfahren nach Anspruch 1,

bei dem die Änderungsvorschrift zumindest teilweise so berechnet wird, dass die Gewichte (w_i) von Eingangsgrößen (x_i), die einen geringen Einfluss auf die Ausgangsgröße (y_i) haben, in Abhängigkeit vom Wert anderer der Gewichte (w_i) verringert werden.

3. Verfahren nach Anspruch 2,

bei dem die Änderungsvorschrift zumindest teilweise so berechnet wird, dass die Gewichte (w_i) von Eingangsgrößen (x_i), die einen geringen Einfluss auf die Ausgangsgröße (y_i) haben, um einen Wert verringert werden, der von der Summe über die Quadrate anderer der Gewichte (w_i) abhängt.

4. Verfahren nach zumindest einem der vorhergehenden Ansprüche,

bei dem die Verringerung der Gewichte (w_i) mit einer vorgebaren Verringerungsrate (λ) erfolgt.

5 5. Verfahren nach zumindest einem der vorhergehenden Ansprü-

che,

bei dem die Abbildung (2, 4, 200-210, 400) mehrere Abbil-
dungsschichten enthält und die Gewichte (w_i) bei der ersten
Abbildungsschicht angeordnet sind.

10 6. Verfahren nach zumindest einem der vorhergehenden Ansprü-
che,

bei dem mindestens eines der Gewichte mit einem Wert, der un-
ter einem Schwellwert liegt, auf Null gesetzt wird.

15 7. Verfahren nach zumindest einem der vorhergehenden Ansprü-
che,

bei dem die Abbildung (2, 4, 200-210, 400) für mindestens ei-
ne der Eingangsgrößen (x_i) eine Übertragungsfunktion auf-
weist, die für betragsmäßig große Werte der mindestens einen
20 der Eingangsgrößen (x_i) eine geringe Steigung aufweist.

8. Verfahren nach zumindest Anspruch 7,

bei dem die Abbildung (2, 4, 200-210, 400) mehrere Abbil-
dungsschichten (2, 4, 200-210, 400) enthält und die Übertra-
25 gungsfunktion in oder nahe der ersten Abbildungsschicht (2,
200-210) angeordnet ist.

9. Verfahren nach zumindest Anspruch 7,

bei dem die Übertragungsfunktion eine sigmoide Übertragungs-
30 funktion ist.

10. Verfahren nach zumindest einem der vorhergehenden Ansprü-
che,

bei dem die Abbildung (2, 4, 200-210, 400) ein neuronales
35 Netz enthält.

11. Verfahren nach zumindest Anspruch 10,

bei dem die Eingangsgrößen (x_i) über künstliche Neuronen mit der mindestens einen durch die Abbildung (2, 4, 200-210, 400) erzeugten Ausgangsgröße (y_i) verknüpft werden.

5 12. Verfahren nach zumindest Anspruch 10,
bei dem die Gewichte (w_i) Gewichte im neuronalen Netz sind.

13. Verfahren nach zumindest einem der vorhergehenden Ansprüche,

10 bei dem die Schritte, dass

- die von der Abbildung (2, 4, 200-210, 400) erzeugte Ausgangsgröße (y_i) mit der von der Abbildung (2, 4, 200-210, 400) zu erzeugenden Ausgangsgröße verglichen wird,
- aus dem Vergleich die Änderungsvorschrift für die Abbildung (2, 4, 200-210, 400) berechnet wird und
- die Änderungsvorschrift zumindest teilweise mit der Maßgabe berechnet wird, dass die Gewichte (w_i) von Eingangsgrößen (x_i) verringert werden, die einen geringen Einfluss auf die Ausgangsgröße (y_i) haben,

20 iterativ wiederholt werden.

14. Verfahren zur Sprachverarbeitung mit Abbildungen (200-210, 400) mehrerer Sätze von Sprachmerkmale enthaltenden Eingangsgrößen (x_i) auf mindestens eine durch eine der Abbildungen (200-210, 400) erzeugten Ausgangsgröße (y_i) und mit mindestens einer durch eine der Abbildungen (200-210, 400) zu erzeugenden Ausgangsgröße, insbesondere nach zumindest einem der Ansprüche 1 bis 13, bei dem

- a) die von den Abbildungen (200-210, 400) erzeugte Ausgangsgröße (y_i) mit der von den Abbildungen (200-210, 400) zu erzeugenden Ausgangsgröße verglichen wird,
- b) aus dem Vergleich Änderungsvorschriften für die Abbildungen (200-210, 400) berechnet werden,
- c) die Änderungsvorschriften so berechnet werden, dass sich dieselben Abbildungen (200-210, 400) für unterschiedliche Sätze von Eingangsgrößen (x_i) ergeben.

15. Verfahren nach zumindest Anspruch 14,
bei dem die Sätze von Eingangsgrößen (x_i) jeweils gleichartige Eingangsgrößen (x_i) enthalten.

5 16. Verfahren nach zumindest Anspruch 15,
bei dem die Sätze von Eingangsgrößen (x_i) durch eine Zeitreihe jeweils gleichartiger Eingangsgrößen (x_i) gebildet werden.

10 17. Verfahren nach zumindest einem der Ansprüche 14 bis 16,
bei dem vorläufige Änderungsvorschriften für die Abbildungen (200-210, 400) jeweils einzelner Sätze von Eingangsgrößen (x_i) berechnet werden und die Änderungsvorschriften aus Mittelwerten der vorläufige Änderungsvorschriften berechnet werden.

15 18. Verfahren nach zumindest einem der Ansprüche 14 bis 17,
bei dem zumindest eine der Abbildungen (200-210, 400) für mindestens eine der Eingangsgrößen (x_i) eine Übertragungsfunktion aufweist, die für betragsmäßig große Werte der mindestens einen der Eingangsgrößen (x_i) eine geringe Steigung aufweist.

20 19. Verfahren nach zumindest Anspruch 18,
bei dem zumindest eine der Abbildungen (200-210, 400) mehrere Abbildungsschichten (200-210, 400) enthält und die Übertragungsfunktion in oder nahe der ersten Abbildungsschicht (200-210) angeordnet ist.

25 20. Verfahren nach zumindest Anspruch 18,
bei dem die Übertragungsfunktion eine sigmoide Übertragungsfunktion ist.

30 21. Verfahren nach zumindest einem der Ansprüche 14 bis 20,
bei dem zumindest eine der Abbildungen (200-210, 400) ein neuronales Netz enthält.

35 22. Verfahren nach zumindest Anspruch 21,

bei dem die Eingangsgrößen (x_i) über künstliche Neuronen mit der mindestens einen durch die mindestens eine der Abbildungen (200-210, 400) erzeugten Ausgangsgröße (y_i) verknüpft sind.

5

23. Verfahren nach zumindest einem der vorhergehenden Ansprüche,

bei dem die Schritte, dass

- die von den Abbildungen (200-210, 400) erzeugte Ausgangsgröße mit der von den Abbildungen (200-210, 400) zu erzeugenden Ausgangsgröße verglichen werden,
- aus dem Vergleich die Änderungsvorschriften für die Abbildungen (200-210, 400) berechnet werden,
- die Änderungsvorschriften so berechnet werden, dass sich dieselben Abbildungen (200-210, 400) für unterschiedliche Sätze von Eingangsgrößen (x_i) ergeben,

iterativ wiederholt werden.

10

15

24. Verfahren nach zumindest einem der Ansprüche 14 bis 23, bei dem die Abbildungen (200-210, 400) mindestens einen gemeinsamen Abbildungsteil (400) aufweisen.

25 25. Verfahren zur Sprachverarbeitung mit einer Abbildung (2, 4, 200-210, 400) von Sprachmerkmale enthaltenden Eingangsgrößen (x_i) auf mindestens eine durch die Abbildung (2, 4, 200-210, 400) erzeugte Ausgangsgröße (y_i), bei dem eine nach einem der Ansprüche 1 bis 24 erzeugte Abbildung (2, 4, 200-210, 400) verwendet wird.

30

26. Anordnung, die eingerichtet ist, ein Verfahren nach zumindest einem der Ansprüche 1 bis 25 auszuführen.

35

27. Computerprogrammprodukt, das Softwarecodeabschnitte enthält, mit denen ein Verfahren nach zumindest einem der Ansprüche 1 bis 25 auf einem Computer ausgeführt werden kann.

Zusammenfassung

Verfahren zur Sprachverarbeitung

5 Die Erfindung betrifft ein Verfahren zur Sprachverarbeitung, bei dem Sprachmerkmale enthaltende Eingangsgrößen auf Ausgangsgrößen abgebildet werden. Bei der Abbildung werden die Eingangsgrößen gewichtet und/oder es werden gleiche Abbildungen für unterschiedliche Sätze von Eingangsgrößen und mindes-
10 tens eine Ausgangsgröße erzeugt.

200017729

2000 E 16711

3ad

21

Figur 1

5

10

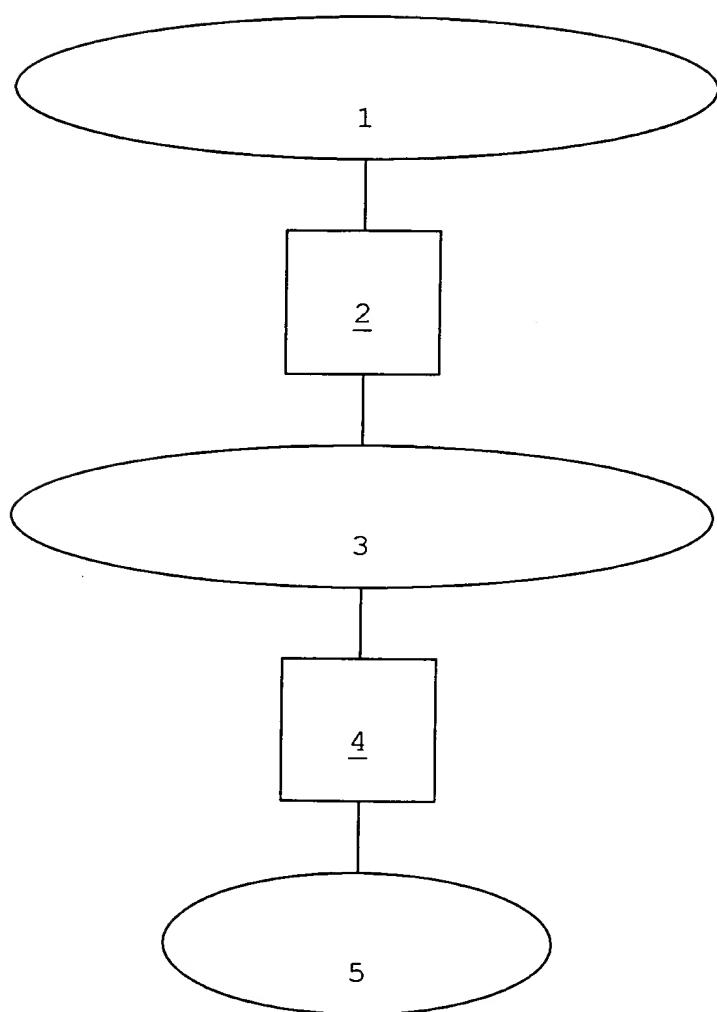
15

20

25

30

35



200017729

2000 E 16711

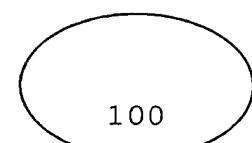
Bad

22

Figur 2

5

10



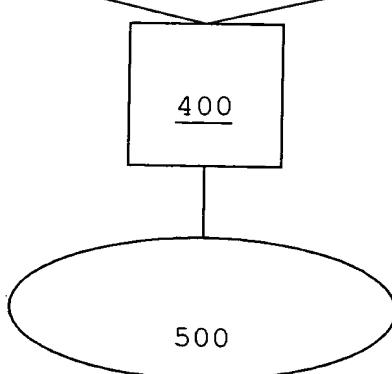
15

20

25

30

35



200017729

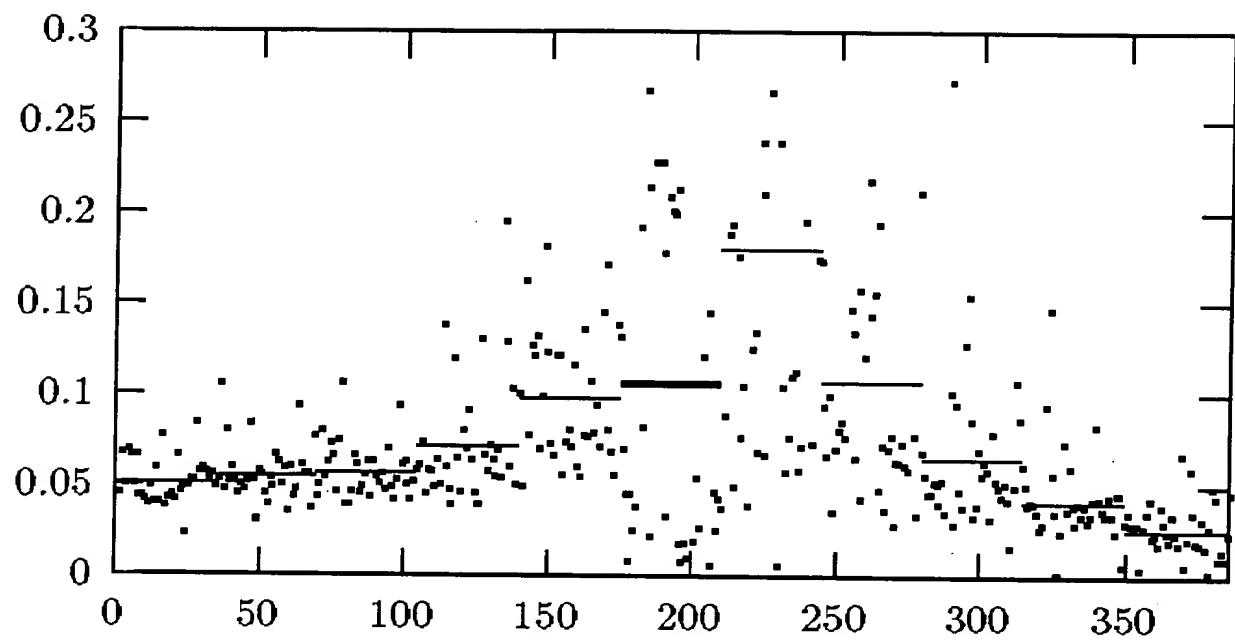
2000 E 16711

Bad

23

Figur 3

5



10

This Page Blank (uspto)